
A tRNA-guided research journey from synthetic chemistry to synthetic biology

DIETER SOLL

Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520, USA

I was first introduced to RNA in 1962. As a newly minted PhD in synthetic organic chemistry I had just joined in Madison, Wisconsin a team of four postdocs charged with the challenging task of chemically synthesizing 64 ribonucleotides by the most modern methods (manual synthesis, of course, requiring gallons of dry pyridine every day). My advisor, Har Gobind Khorana, a man of boundless energy, vision and courage, was convinced that these RNA triplets would be essential for the elucidation of the genetic code. It was an exciting time in molecular biology as the first details of the mechanism of protein synthesis and the genetic code emerged. Then, after three years of intense work with great team spirit the code was cracked! At the end of my time in Wisconsin I had become a molecular biologist, spellbound by the genetic code and by tRNA. I would never lose interest in this molecule, and it guided my research for the next five decades.

Fast forward to the mid-nineteen-nineties, when the *RNA* journal was launched, the era of genome sequencing began, and I returned to studies of the genetic code. At that time we dabbled for the first time in using Archaea for our research; we discovered that tRNA-dependent asparagine (Asn) formation provides Asn-tRNA for protein synthesis, and later showed that in many organisms this process is also the required supply route for the free amino acid. Then a phone call changed my research direction for the next 15 years. Carl Woese, whose early inquiries into the genetic code and tRNA, summarized in his influential “little book” *The Genetic Code, the Molecular Basis for Genetic Expression* (published in 1967; now available used from Amazon as a collector’s item for \$604!), had made a big impression on me back then. Carl had just called. He told me that the genome sequence of the first archaeon, *Methanococcus jannaschii* (now renamed *Methanocaldococcus jannaschii*), would appear in print soon, and that the genome lacked the gene for the essential cysteinyl-tRNA synthetase. “How are you going to explain this?” he asked. Without

any hesitation I replied that I would solve the riddle, totally unaware that it would take us a decade of sustained and unfamiliar work with many archaeal species to come up with the correct and exciting solution! Thus, we launched investigations with organisms whose names were hard to pronounce, even more difficult to grow, that were anaerobes, and some of them with stunning optimal growth temperatures exceeding the boiling point of water; all of this was totally unfamiliar to us. However, given our thrilling unexpected first results and considering the vast diversity of microbes that would soon be available via their genome sequences, we felt that careful analyses of these organisms for anything that diverged from the accepted view of protein synthesis and genetic coding would possibly lead to surprising new concepts. And given the vast organismal diversity out there, we wondered how many different routes (or exceptions to the currently accepted dogma) of protein synthesis, tRNA formation and the genetic code would be represented in this biological universe.

Here are some of the surprises. (1) *Nanoarchaeum equitans*, the tiny archaeal parasite, showed us that tRNA biosynthesis need not involve a complete gene, but that intact and active tRNAs can be assembled from small pieces of tRNA, a fact that later turned out to be true for several organisms involving different sizes and numbers of pieces. (2) *Nanoarchaeum* also revealed to us that here could be life without the essential RNaseP, the enzyme that processes the 5'-terminus of mature tRNA from the customary precursor molecule. However, this tiny organism that underwent massive genome reduction during its evolution, had spatially arranged its tRNA genes in the genome so that transcription started at the 5'-terminus of the mature tRNA species! (3) *Methanopyrus kandleri*, a hyperthermophile growing at 110°C, gave us another surprise as the biosynthesis of 88% of its tRNA genes requires C→U RNA editing. And the discovery of this hyperthermophile cytidine deaminase quickly led to a crystal structure of a complete RNA editing enzyme! (4) The work on *Nanoarchaeum* tRNA biosynthesis brought us

Corresponding author: dieter.soll@yale.edu

Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.050625.115>. Freely available online through the *RNA* Open Access option.

© 2015 Soll This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

back to the elusive RNA ligase involved in mammalian tRNA maturation. The HeLa cell 3'-P RNA ligase activity was biochemically detected about 30 years earlier, but turned out to be too labile to be purified. Fortunately, *Methanopyrus kandleri* came to our rescue as it allowed complete purification of a minute amount of protein with 3'-P RNA ligase activity, identified as RtcB with sequencing. The mammalian homolog had earlier been detected in the U2 spliceosome by the Reinhard Lührmann lab, but its activity remained unknown. Together with the Lührmann and Javier Martinez labs we first characterized this essential heterotetrameric mammalian 3'-P RNA ligase, which is now known to be responsible for not only tRNA ligation but also the unfolded protein response. (5) Pyrrolysine (Pyl). The discovery of a 22nd genetically encoded amino acid (Joseph Krzycki & Michael Chan labs, 2002) in *Methanosarcina* was exciting, and led us to identify Pyl-tRNA synthetase (PylRS) as the first tRNA synthetase specific only for a non-canonical amino acid (ncAA). PylRS differed from most other tRNA synthetases as it does not recognize the tRNA anticodon as identity element, and that the PylRS:tRNA^{Pyl} complex shows an intricate specialized synthetase:tRNA interaction surface. As such it would be an ideal orthogonal pair for genetic code expansion. (6) Studies with archaea were essential in the definition of tRNA-dependent amino acid transformations that provide routes to form Gln-tRNA and Asn-tRNA in many organisms in all three kingdoms. (7) Next, the ultimate answer to Carl Woese's question about the missing cysteinyl-tRNA synthetase: we discovered that methanogens form Cys-tRNA through tRNA-dependent transformation of phosphoserine (Sep), which was attached to tRNA by phosphoserine-tRNA synthetase (SepRS), the second tRNA synthetase specific only for a ncAA. (8) The finding of a Sep-tRNA^{Cys} intermediate in Cys-tRNA formation focused our imagination immediately on a possible Sep-tRNA^{Sec} intermediate in the unknown selenocysteine (Sec) biosynthesis pathway of archaea and eukaryotes. And yes, we demonstrated the existence of this intermediate, and converted it to Sec-tRNA^{Sec} by a protein of unknown biological function described 14 years earlier by Erik Sontheimer! This work then led to the firm conclusion that tRNA-dependent amino acid transformations synthesize Asn-tRNA, Gln-tRNA, Cys-tRNA and Sec-tRNA in many organisms. Furthermore, Sec is the only genetically encoded amino acid that lacks a cognate aminoacyl-tRNA synthetase. (9) Genetic Code variations. This is a most exciting field that combines classical protein sequence data (not derived from DNA sequence), biochemical work with tRNAs, and protein and tRNA sequences derived from metagenomic, genomic, and single cell genomic sequence data. We identified (together with Franz Lang) recoding events in yeast and other fungal mitochondria, and the lineage of the recoding tRNA; found that *Acetohalobium arabaticum* expands its genetic code (from 20 to 21 amino acids to include Pyl) depending on the carbon source of the growth medium; and that SR1 bacteria use UGA as a 5th glycine co-

don, thus making its DNA "useless" for horizontal gene transfer (in collaboration with the Mircea Podar and Tanja Woyke labs).

In 2008 another development changed again the course of my research. The question as to how the genetic code evolved was a lively topic at the time the code was cracked. Obviously, genetic code alterations change the meaning of codons that lead to mistranslation and inaccuracies in the proteome. While this may have been acceptable at an early stage of code development, at the stage of the present code "no new amino acid could be introduced without disrupting too many proteins" (Crick 1968). Thus Crick considered the present code the result of a "frozen accident" unable to evolve further even if the current state were suboptimal. Therefore, the exquisite substrate specificity of aminoacyl-tRNA synthetases was an unchallenged dogma, and I had such a statement prominently in all my grant applications! I was confounded when we discovered (in 2008) that *E. coli* was able to accommodate about 10% mismade proteins without much negative impact. Realizing this fact, I then decided to become a synthetic biologist and enter the field of genetic code expansion. What is required for this? To take a non-canonical amino acid to make ncAA-tRNA specific for a certain codon, deliver it to the ribosome, make sure that the mRNA:tRNA coding interactions are productive, and that the ribosome can handle the unusual geometry of a ncAA. Well, these are all experiments for someone trained in working with tRNA, aminoacyl-tRNA synthetases, and ribosomal protein synthesis. Thus, I thought we ought to be able to do it!

For introduction of ncAAs "orthogonal" components of the translation machinery were needed, foremost aminoacyl-tRNA synthetases, tRNAs and elongation factors. Once again it turned out that the archaeal kingdom already had perfect orthogonal aminoacyl-tRNA synthetase:tRNA pairs, e.g., SepRS and PylRS. And so we joined the dynamic field of genetic code expansion with the strategy of deliberately attempting to circumvent protein synthesis quality control by mutagenizing tRNA synthetases and elongation factors and designing synthetic tRNAs. This has led robust methods of site-specific insertion, programmed by UAG, of Sep or Sec into any desired protein in *E. coli*. And "playing" with the natural *E. coli* Sec insertion machinery we found it capable of incorporating Sec directed by 60 of the 64 codons! Now the challenge is multiple sense codon recoding!

Our work taught me (and my group) another facet of scientific inquiry: how to find a journal interested in publishing the results of tRNA research. Strong competition in the tRNA field and a seeming lack of appreciation for nucleic acid topics by some journals led me (jointly with Richard Walker and Albert Jones) to launch the journal *Nucleic Acids Research* in 1974. Together with the *RNA* journal they have proved welcoming homes for tRNA research. Moreover, tRNA, exciting as ever, is still of interest to the "CNS" journals (to borrow a phrase from Harry Noller, this issue!)

tRNA is a marvelous molecule in another sense: it inspired many excellent young minds to join our Yale team as students and postdocs whose skills often surpassed mine (exemplified by the beautifully written tribute to tRNA by Michael Ibba, this issue); without their contributions the pages above would be blank.

Looking to the future: organismal variants with different codes, coding based on mRNA containing six or more different bases, ribosomes with altered specificities translating α -, β -, and γ -amino acids for production of biomaterials; a tRNA biologist's dream of a brave new synthetic world.